DYNAMIC SPECTRUM CLASSIFICATION BY DIVERGENCE-BASED KERNEL MACHINES AND ITS APPLICATION TO THE DETECTION OF WORN-OUT BANKNOTES

Tsukasa Ishigaki

Japan Science and Technology Agency
4-6-7 Minami Azabu, Minato-ku
Tokyo 106-8569, Japan
ishigaki@ism.ac.jp

Tomoyuki Higuchi

Institute of Statistical Mathematics
4-6-7 Minami Azabu, Minato-ku
Tokyo 106-8569, Japan
higuchi@ism.ac.jp

ABSTRACT

In the kernel method, the appropriate selection or design of the kernel function is important for the construction of a high-performance classifier. The present paper describes a dynamic spectrum classification method using kernel classifiers with the divergence-based kernel and its application to the detection of worn-out banknotes. We introduce the divergence-based kernel that was proposed as a measure between two probability distributions into the dynamic spectrum classification. The present method is applied to the detection of worn-out banknotes by using acoustic signals for the facilitation of identifying counterfeit banknotes. As a result, the classification performance using the divergence-based kernel is shown to have better performance than those using common kernels such as the Gaussian kernel or the polynomial kernel.

***Index Terms*—** kernel method, pattern recognition, acoustic signal processing, spectrum classification, acoustic applications

1. INTRODUCTION

Kernel methods have been applied successfully in various fields. In the classification problem using the kernel classifier, it is significant to properly select or design the kernel function. A number of studies have examined kernel design, and it is well known that the performance of the kernel classifier changes significantly depending on the kernel function. However, there is no criterion that provides selection or design methods of the kernel function to construct a high-performance classifier. The constructed classifier might become a low-performance classifier if the selection of the kernel function is not appropriate for the data distribution. In contrast, a high-performance classifier can be obtained by finding an appropriate kernel for a specific problem.

In the present paper, we introduce the divergence-based kernel into the normalized dynamic spectrum classification and apply this method to the detection of worn-out banknotes. The divergence gives a measure between two probability distributions, and the divergence-based kernel was proposed as a measure between two probability distributions in the kernel method. We apply the divergence-based kernel to the classification of dynamic spectra, which is often used for speech recognition or earthquake analysis, by using Support Vector Machine (SVM) [1] and Relevance Vector Machine (RVM) [2]. The authors previously demonstrated a high-accuracy classification method for the normalized spectrum by using Kullback-Leibler divergence-based kernel and applied this method to fault detection of an LP gas instrument [3]. The obtained result showed a higher classification performance than some common methods. In the present paper, we show that the method can be efficiently applied to the dynamic spectrum classification problem. The classification method is efficient for the detection of worn-out banknotes for the facilitation of identifying counterfeit banknotes or preventing paper jams in automated teller machines (ATMs).

2. DIVERGENCE-BASED KERNEL AND KERNEL CLASSIFIER

2.1. Divergence-based kernel

Kullback-Leibler (KL) divergence, which is the most famous divergence, is used in the present study. The continuous form of the KL divergence between $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$ with random variable $\boldsymbol{x}$ is defined as follows:

$$D_{KL}(p(\boldsymbol{x}), q(\boldsymbol{x})) = \int_{-\infty}^{\infty} p(\boldsymbol{x}) \log \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} d\boldsymbol{x}. \quad (1)$$

The divergence is introduced to kernel function that measures the similarity of two features in the present classification method. The kernel function must be a symmetric function. However, the KL divergence is not symmetric, and the symmetric KL divergence is considered to be

$$SD_{KL}(p(\boldsymbol{x}), q(\boldsymbol{x})) = \int_{-\infty}^{\infty} \{p(\boldsymbol{x}) - q(\boldsymbol{x})\} \log \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} d\boldsymbol{x} \quad (2)$$

and the KL kernel is defined as

$$K_{KL}(p(\boldsymbol{x}), q(\boldsymbol{x})) = \exp\left[-\rho SD\left\{p(\boldsymbol{x}), q(\boldsymbol{x})\right\}\right], \quad (3)$$

where ρ is a constant.

The KL kernel is usually recognized as a kernel that measures the similarity of two probability distributions in speech and image classification applications [4]. However, in the present study, we do not consider the probability density because we deal not with the probability function but with the frequency spectrum. The feature vector consists of the discretized frequency spectrum, as $\boldsymbol{x}_i = [x_i(1) \cdots x_i(M)]^T$, where M is the number of elements of the feature vector. In the present applications, the KL kernel is defined as follows:

$$K_{KL}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left\{-\rho SD(\boldsymbol{x}_i, \boldsymbol{x}_j)\right\}, \quad (4)$$

$$SD(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sum_{k=1}^{M} \left[\left\{x_i(k) - x_j(k)\right\} \log \frac{x_i(k)}{x_j(k)}\right]. \quad (5)$$

Note that the input feature vector to the divergence-based kernel is no longer required to be the probability distribution function. Equations (4) and (5) can deal with a real valued vector if all of the elements of the feature vector are positive and the feature vector satisfies the following condition:

$$\sum_{k=1}^{M} \boldsymbol{x}_i(k) = \sum_{k=1}^{M} \boldsymbol{x}_j(k). \quad (6)$$

2.2. Support Vector Machine and Relevance Vector Machine

Here, we employ the SVM, which is a popular kernel machine, and the RVM, which is a Bayesian kernel machine. The functional form of the output of the two kernel machines is identical, and both machine are well-known as sparse classifiers.

Over the past ten years, the combination of the SVM and the kernel method that realizes a nonlinear classification has been successfully implemented in various fields. The optimization procedure of SVM converges on theoretical analysis if the Gram matrix of the kernel function is positive semi-definite because the optimization form of the SVM is a convex quadratic program. However, the convergence of optimization procedure of the SVM with a divergence-based kernel is not ensured because the Gram matrix of most divergence-based kernels are non-positive semi-definite. On the other hand, the convergence of the optimization of the RVM is ensured, even if the Gram matrix of the kernel functions are non-positive definite [2], and the RVM tends to be sparser than the SVM.

3. DETECTION OF FATIGUED BANK NOTES BY ACOUSTIC SIGNAL

3.1. Maintenance of Paper Quality of Bank Notes

In some countries, the quality of paper banknotes in circulation is maintained at a high level. The main reasons for this are (1) to facilitate the detection of counterfeit banknotes, and (2) to prevent paper jam problems in automated teller machines (ATMs). Old, stained, or badly wrinkled banknotes (referred to hereinafter as worn-out banknotes) frequently cause these problems. Such banknotes are removed from circulation and discarded by the central bank. For this procedure, automated detection of worn-out banknotes or ATMs having this capability have been used. These machines perform detection by sensing the optical or acoustic [5] properties of banknotes. The optical detection method using images or infrared sensors has primarily been employed for practical applications. However, this method requires mechanisms for the alignment of the direction and position of banknotes or discrimination of the type of banknote. On the other hand, detection by acoustic signal does not require such mechanisms (only a common microphone and a signal processor, both of which are inexpensive, are needed). Therefore, the implementation of a worn-out banknote detection machine based on acoustic analysis in existing ATMs is easier than the implementation of optical analysis methods. Moreover, optical methods for detecting worn-out banknotes examine only certain areas of banknotes that are determined empirically to be likely to have wrinkles or stains. Therefore, optical methods are difficult to apply directly to other applications such as detection of worn-out banknotes of various countries or other discriminant techniques of paper quality without empirical knowledge in specific fields. In contrast, the acoustic method can detect differences in sound depending on the paper condition or quality without empirical or prior knowledge. Therefore, application of the acoustic method to other techniques is comparatively easy. However, the performance of worn-out banknote detection by conventional acoustic methods is not sufficient (approximately 90%).

In this section, the dynamic spectrum classification method based on the SVM and the RVM with the divergence-based kernel is applied to the worn-out banknote detection problem. Here, two data sets are used for verification of classification performance, and the sound of flicking of banknotes is considered as the analyzed object. Normalization of the power spectrum of the obtained signal is natural in this case because the characteristics of paper quality vary with the frequency of the signal rather than its amplitude.

3.2. Measurement of Two Types of Acoustic Signals of Bank Notes

The sound of 'flicking' banknotes was measured by two methods. In the first method, a fixed banknote on a platform is

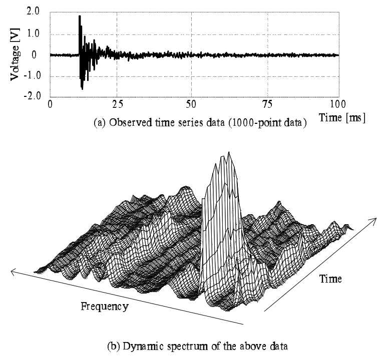

Fig. 1. Observed impact acoustic signal and estimated dynamic spectrum

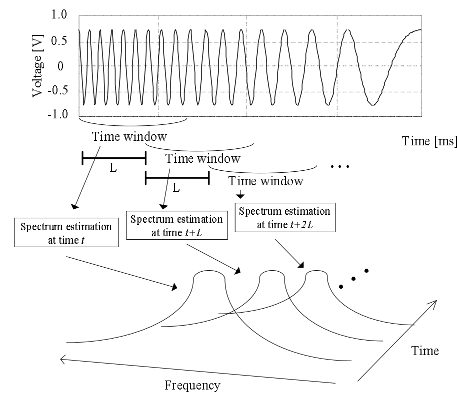

Fig. 2. Schematic diagram of dynamic spectrum estimation with time window and time lag

flicked by plastic bar. In this experiment (experiment 1), the sounds of 50 reusable banknotes and the sounds of 50 worn-out banknotes were measured from three sheets of reusable banknotes and three sheets of artificially-aged banknotes, respectively. The data set measured by this method is hereinafter referred to as the fixed banknote data set.

In another experiment (experiment 2), 10 people were asked to hold a banknote in one hand and 'flick' the banknote with a finger of the other under various flicking conditions, which included how the note was held, the position with which the note was held, the strength of the flick, and the area of the banknote that was flicked. If good performance can be obtained using this data set and the present classification method, then the classification method may be applicable to different types of ATM. Data for 100 reusable banknotes and 100 worn-out banknotes were obtained through 10 trials for each type of bank note and each person. The data set obtained by this measurement is referred to herein as the freely flicked data set.

Both types of data sets were measured at a sampling frequency of 10 kHz by a microphone placed approximately 100 mm from the banknotes. Observed time series data in the freely flicked data set and the dynamic spectrum estimated from time-varying AR coefficients smoothed by the first-order trend model are shown in Figure 1. When the dynamic spectrum is estimated, it is impractical to calculate the spectrum at all times t due to the associated computational cost. The spectrum was estimated at time t for time lag L using the time window including the length of time series T_w data, as shown in Figure 2. Figure 1(b) was estimated by setting $T_w = 100$ and $L = 10$.

3.3. Feature Vector

A total of 1,000 data points $z_{it}(t = 1, 2, \cdots, T)$ were extracted from the data i. The starting point of the signal of the 1,000 points is determined at the point in time at which the absolute observed value exceeds 0.1 volts. The feature vector $\boldsymbol{x}_i$ is then constructed, and these features are used for classification. Hereafter, the time window T_w and time lag L are set to 100 and 30, respectively. These values are determined experimentally.

Some features such as the stationary periodogram, the stationary spectrum estimated by the AR method, the dynamic periodogram, and the smoothed dynamic periodogram are employed for classification of the banknote data set. Here, we deal with the dynamic spectrum estimated by the time-varying AR model because the employment of this feature showed the best classification performance among the selected features.

The dynamic spectrum s_{it} of datum i at time t with time lag L can be obtained by the following equation:

$$s_{it}(k) = \frac{\sigma_{it}^2}{\left|1 - \sum_{j=1}^{K_{it}} a_{it}(j) e^{-jk\sqrt{-1}}\right|^2}, \tag{7}$$

where K_{it} is the optimal AR order in terms of the AIC at time t. The AR coefficient $a_{it}(j)$ and σ_{it}^2 are estimated by the Yule-Walker method at time t. The dynamic spectrum components at time t from 0 Hz to 5 kHz in 100-Hz increments were selected as a feature that is normalized at each time as follows:

$$\bar{s}_{it} = \frac{1}{\sum_{k=1}^{N_f} s_{it}(k)} [s_{it}(1), \cdots, s_{it}(N_f)]^T. \tag{8}$$

Here, $N_f = 51$. The dynamic spectrum is then constructed as

$$\boldsymbol{x}_i = [\bar{s}_{i1}^T, \bar{s}_{iL}^T, \bar{s}_{i2L}^T, \cdots, \bar{s}_{iN_sL}^T]^T, \tag{9}$$

where N_s is the number of calculated spectra $\bar{s}_{it}$.

3.4. Experiments and Results

Half of the data extracted for reusable banknotes and half of the data extracted for worn-out banknotes were selected randomly from each data set and were assigned as a training data

Table 1. Correct answer percentages and standard deviations of the detection of worn-out banknotes by combinations of kernel functions and kernel classifier

	Experiment 1		Experiment 2	
	SVM	RVM	SVM	RVM
1st-order polynomial	90.1 (3.9)	87.5 (4.2)	84.5 (3.5)	84.7 (3.8)
2nd-order polynomial	92.2 (3.5)	88.1 (4.9)	86.2 (3.4)	84.6 (3.6)
3rd-order polynomial	92.1 (3.2)	88.4 (4.5)	85.5 (3.3)	85.0 (3.4)
4th-order polynomial	92.3 (3.4)	88.5 (4.6)	85.6 (3.0)	85.0 (3.6)
5th-order polynomial	92.7 (3.2)	88.0 (4.6)	86.0 (3.1)	84.9 (4.0)
Sigmoid	90.0 (4.3)	85.9 (4.5)	82.7 (3.9)	82.1 (3.7)
Gaussian	93.4 (3.6)	88.9 (3.2)	87.6 (3.1)	86.6 (3.2)
Laplacian	98.1 (2.3)	95.5 (2.8)	87.8 (3.6)	85.6 (3.4)
Sublinear	97.7 (3.5)	96.6 (3.1)	89.6 (3.4)	88.9 (3.2)
χ^2	98.6 (2.1)	95.6 (2.6)	90.9 (2.7)	88.1 (2.7)
Hellinger	99.1 (1.8)	96.6 (2.2)	91.6 (2.4)	90.0 (2.5)
Bhattacharyya	**99.4** (1.3)	95.7 (2.1)	91.8 (2.5)	89.8 (2.2)
Kullback-Leibler	**99.4** (1.3)	96.0 (2.0)	**92.0** (2.3)	90.1 (2.3)

set in each experiment. The remaining data were then used for validation in both experiments. This operation was repeated 100 times. The parameter with respect to the regularization of SVM and RVM and the kernel function is determined by 10 two-fold cross-validation tests. The 13 kernel functions were used in the experiments. The first- to fifth-order polynomial, Sigmoid, and Gaussian kernels are used in various situations. The Laplacian, Sublinear, and χ^2 kernels are often used for histogram classification. The Hellinger divergence-based kernel, the Bhattacharyya divergence-based kernel, and the KL divergence-based kernel were proposed in order to treat the probability distributions on the kernel method. Dynamic spectra were created as the time window of $T_w = 100$ and $L = 30$. Table 1 shows the average accuracy percentages and the standard deviations in the two experiments with respect to the combinations of kernel classifier and kernels functions in the 100 operations.

3.5. Discussion

The experiment using the fixed banknote data set and the Bhattacharyya divergence-based kernel and the KL divergence-based kernel had correct answer rates of nearly 100%. In the experiment using the freely flicked data set, the correct answer rates obtained using the Bhattacharyya kernel and the KL kernel had correct answer rates that consistently approached 92%. The use of the divergence-based kernel provided high accuracy rates compared to the use of common kernels. These classification results are much improved compared to the existing results. The results revealed that the present method is an effective method for detecting worn-out banknotes.

Generally, the classification performances of the SVM have exceeded those of the RVM by approximately 3%. The divergence-based kernels are non-positive definite, however, the optimization procedure converges successfully in all cases of the present study. Therefore, the divergence-based kernel is considered to be sufficiently practical in the present problem.

4. CONCLUSION

In the present paper, we described the dynamic spectrum classification with the divergence-based kernel and the detection of the work-out banknote. The SVM and RVM were applied to the detection task and the high classification performance realized by the proposed method. The results depended on the kernel selection, which is significant for further advancement of classification performance as well as feature selection or choice of classifier. In the future, the influence of the fatigue level of banknotes on practical implementation of the proposed method will be examined.

5. REFERENCES

[1] N. Cristianini and J. S. Taylor, *An Introduction to support vector machines and other kernel-based learning methods*, Cambridge University Press, 2000.

[2] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res*,vol. 1, pp. 211-244, 2001

[3] T. Ishigaki, T. Higuchi and K. Watanabe, "Spectrum classification for early fault diagnosis of LP gas pressure regulator based on Kullback-Leibler kernel," *IEEE MLSP'06.*, pp.453–458, 2006.

[4] P. J. Moreno, P. P. Ho and N. Vasconcelos, "A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications," *NIPS'03.*, 2003.

[5] M. Teranishi, T. Matsui, S. Omatsu and T. Kosaka, "Neuro-classification of fatigued bill based on tensional acoustic signal", *Proc. IEEE SMCia'05*, pp. 173-177, 2005